

Orthogonality and probability: mixing times

Yevgeniy Kovchegov*

Abstract

We produce the first example of bounding total variation distance to stationarity and estimating mixing times via orthogonal polynomials diagonalization of discrete reversible Markov chains, the Karlin-McGregor approach.

1 Introduction

If P is a reversible Markov chain over a sample space Ω , and π is a reversibility function (not necessarily a probability distribution), then P is a self-adjoint operator in $\ell^2(\pi)$, the space generated by the inner product

$$\langle f, g \rangle_\pi = \sum_{x \in S} f(x)g(x)\pi(x)$$

induced by π . If P is tridiagonal operator (i.e. a nearest-neighbor random walk) on $\Omega = \{0, 1, 2, \dots\}$, then it must have a simple spectrum, and is diagonalizable via orthogonal polynomials as it was studied in the 50's and 60's by Karlin and McGregor, see [2], [8]. There the extended eigenfunctions $Q_j(\lambda)$ ($Q_0 \equiv 1$) are orthogonal polynomials with respect to a probability measure ψ and

$$p_t(i, j) = \pi_j \int_{-1}^1 \lambda^t Q_i(\lambda) Q_j(\lambda) d\psi(\lambda) \quad \forall i, j \in \Omega,$$

where π_j ($\pi_0 = 1$) is the reversibility measure of P .

In this paper we are testing a possibility of calculating mixing rates using Karlin-McGregor diagonalization with orthogonal polynomials. In order to measure the rate of convergence to a stationary distribution, the following distance is used.

Definition 1. If μ and ν are two probability distributions over a sample space Ω , then the *total variation distance* is

$$\|\nu - \mu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\nu(x) - \mu(x)| = \sup_{A \subset \Omega} |\nu(A) - \mu(A)|$$

Observe that the total variation distance measures the coincidence between the distributions on a scale from zero to one.

*Department of Mathematics, Oregon State University, Corvallis, OR 97331-4605, USA
kovchegy@math.oregonstate.edu

If $\rho = \sum_{k=0}^{\infty} \pi_k < \infty$, then $\nu = \frac{1}{\rho}\pi$ is the stationary probability distribution. If in addition, the aperiodic nearest neighbor Markov chain originates at site i , then the total variation distance between the distribution $\mu_t = \mu_0 P^t$ and ν is given by

$$\|\nu - \mu_t\|_{TV} = \frac{1}{2} \sum_j \pi_j \left| \int_{(-1,1)} \lambda^t Q_i(\lambda) Q_j(\lambda) d\psi(\lambda) \right|,$$

as measure ψ contains a point mass of weight $\frac{1}{\rho}$ at 1, see [3].

The rates of convergence are quantified via mixing times. In the case of a Markov chain over an infinite state space with a unique stationary distribution, the notion of a mixing time depends on the state of origination of the chain.

Definition 2. Suppose P is a Markov chain with a stationary probability distribution ν that commences at $X_0 = i$. Given an $\epsilon > 0$, the mixing time $t_{mix}(\epsilon)$ is defined as

$$t_{mix}(\epsilon) = \min \{t : \|\nu - \mu_t\|_{TV} \leq \epsilon\}$$

In the case of a nearest-neighbor process on $\Omega = \{0, 1, 2, \dots\}$ commencing at i , the corresponding mixing time has the following simple expression in orthogonal polynomials

$$t_{mix}(\epsilon) = \min \left\{ t : \sum_j \pi_j \left| \int_{(-1,1)} \lambda^t Q_i(\lambda) Q_j(\lambda) d\psi(\lambda) \right| \leq 2\epsilon \right\}$$

Observe that the above expression is simplified when $i = 0$. Here we concentrate on calculating mixing times for simple positive recurrent nearest-neighbor Markov chains over Ω , originating from $i = 0$. Our main result concerns the distance to stationarity for a simple random walk with a drift. In the main theorem and its corollary we will explore the following Markov chain

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ q & r & p & 0 & \dots \\ 0 & q & r & p & \ddots \\ 0 & 0 & q & r & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix} \quad q > p, \quad r > 0$$

Theorem 1. Suppose the above Markov chain begins at the origin, $i = 0$. Consider the orthogonal polynomials Q_n for the chain. Then the integral $\int_{(-1,1)} \lambda^t Q_n(\lambda) d\psi(\lambda)$ can be expressed as

$$\frac{(1+q-p)(q+r)-q}{(1+q-p)(q+r)} \cdot \left(-\frac{q}{q+r}\right)^{t+n} + \left(\sqrt{\frac{q}{p}}\right)^n \left(\frac{p}{q+r}\right) \frac{1}{2\pi i} \oint_{|z|=1} \frac{(\sqrt{pq}(z+z^{-1})+r)^t z^n (z-z^{-1})}{\left(z - \sqrt{\frac{p}{q}} \frac{r+(1+q-p)}{2(q+r)}\right) \left(z - \sqrt{\frac{p}{q}} \frac{r-(1+q-p)}{2(q+r)}\right)} dz$$

and the total variation distance $\|\nu - \mu_t\|_{TV}$ is bounded above by

$$A \left(\frac{q}{q+r}\right)^t + B(r + 2\sqrt{pq})^t,$$

where $A = \frac{(1+q-p)(q+r)-q}{(1+q-p)(1-2p)}$ and $B = \frac{\left(\frac{p}{q+r}\right)\left(1+\frac{1}{\sqrt{pq}-p}\right)}{\left(1-\sqrt{\frac{p}{q}}\frac{r+(1+q-p)}{2(q+r)}\right)\left(1+\sqrt{\frac{p}{q}}\frac{r-(1+q-p)}{2(q+r)}\right)}$. Therefore, taking $\varepsilon \downarrow 0$, the mixing time

$$t_{mix}(\varepsilon) = O\left(\frac{\log(\varepsilon)}{\log m(p, q)}\right),$$

where $m(p, q) = \max\left[\left(r + 2\sqrt{pq}\right), \left(\frac{q}{q+r}\right)\right]$.

Observe that in the above complex integral all three finite poles are located inside the unit circle. Thus we only need to consider a pole at infinity.

The proof is provided in section 3. The result in Theorem 1 is the first instance the Karlin-McGregor orthogonal polynomials approach is used to estimate mixing rates. As it was suggested in [3] we would like the approach to work for a larger class of reversible Markov chains over an infinite state space with a unique stationary distribution. There is an immediate corollary (see section 3):

Corollary. If $\frac{q}{q+r} > r + 2\sqrt{pq}$,

$$\|\nu - \mu_t\|_{TV} \geq A\left(\frac{q}{q+r}\right)^t - B(r + 2\sqrt{pq})^t$$

for t large enough, i.e. we have a lower bound of matching order.

Observe that one can easily adjust these results for any origination site $X_0 = i$. In the next section we will compare the above Karlin-McGregor approach to some of the classical techniques for estimating the “distance to stationarity” $\|\nu - \mu_t\|_{TV}$.

2 Comparison to the other techniques

For the case of geometrically ergodic Markov chains, there are several techniques that produce an upper bound on the distance to stationarity that were developed specifically for the cases when the sample space is large, but finite. These methods are not directly applicable to chains on general state spaces. The coupling method stands out as the most universal. Here we compare the geometric rate in Theorem 1 to the one obtained via a classical coupling argument. Then we explain why other geometric ergodicity methods based on renewal theory will not do better than coupling. See [6] and [4] for detailed overview of geometric convergence and coupling.

2.1 Geometric convergence via coupling

Consider a coupling process (X_t, Y_t) , where $X_0 = 0$ as in Theorem 1, while Y_0 is distributed according to the stationary distribution $\nu = \frac{1}{\rho}\pi$. A classical Markovian coupling construction allows X_t and Y_t evolve independently until the coupling time $\tau_{coupling} = \min\{t : X_t = Y_t\}$. It is natural to compare $P(\tau_{coupling} > t)$ to $P(\tau > t)$, where $\tau = \min\{t : Y_t = 0\}$ is a hitting time, as the chain is positive recurrent.

Now, simple combinatorics implies, for $k \geq n$,

$$P(\tau = k \mid Y_0 = n) = \sum_{i,j: 2i+j=k-n} \frac{k!}{i!(i+n)!j!} p^i q^{i+n} r^j$$

Therefore

$$P(\tau > t) \leq \frac{1}{p\rho} \sum_{k: k>t} \left(\sum_{i,j,n: 2i+j=k-n} \frac{k!}{i!(i+n)!j!} p^{i+n} q^i r^j \right),$$

where ‘ \leq ’ appears because $\pi_0 = 1 < \frac{1}{p}$, but it does not change the asymptotic rate of convergence, i.e. we could write ‘ \approx ’ instead of ‘ \leq ’. The right hand side can be rewritten as

$$\frac{1}{p\rho} \sum_{k: k>t} \sum_{j=0}^k \binom{k}{j} r^j (p+q)^{k-j} \ell(k-j)$$

for $\ell(m) = P(Y \geq m/2)$, where Y is a binomial random variable with parameters $(m, \tilde{p} = \frac{p}{p+q})$. Now, by Cramér’s theorem, $\ell(m) \sim e^{[\log 2 + \frac{1}{2} \log \tilde{p} + \frac{1}{2} \log(1-\tilde{p})]m}$, and therefore

$$P(\tau > t) \sim \frac{1}{p\rho} \sum_{k: k>t} (r + 2\sqrt{pq})^k = \frac{r + 2\sqrt{pq}}{p\rho(\sqrt{q} - \sqrt{p})^2} (r + 2\sqrt{pq})^t \quad (1)$$

Recall that in the Corollary, if q is sufficiently larger than r and p , then $\left(\frac{q}{q+r}\right)^t$ dominates $(r + 2\sqrt{pq})^t$, and the total variation distance

$$\|\nu - \mu_t\|_{TV} = A \left(\frac{q}{q+r}\right)^t \pm B(r + 2\sqrt{pq})^t,$$

where A and B are given in Theorem 1 of this paper. Thus we need to explain why, when q is sufficiently large, in the equation (1), we fail to notice the dominating term of $\left(\frac{q}{q+r}\right)^t$. In order to understand why, observe that the second largest eigenvalue $\left(-\frac{q}{q+r}\right)$ originates from the difference between $\tau_{coupling}$ and τ . In fact, Y_t can reach state zero without ever sharing a site with X_t (they will cross each other, of course). Consider the case when p is either zero, or close to zero. There, the problem reduces to essentially coupling a two state Markov chain with transition probabilities $p(0,1) = 1$ and $p(1,0) = \frac{q}{q+r}$. Thus the coupling time will be expressed via a geometric random variable with the failure probability of $\frac{q}{q+r}$.

Of course, one could make the Markov chain P “lazier” by increasing r at the expense of p and q , while keeping proportion $\frac{q}{p}$ fixed, i.e. we can consider $P_\varepsilon = \frac{1}{1+\varepsilon}(P + \varepsilon I)$. This will minimize the chance of X_t and Y_t missing each other, but this also means increasing $(r + 2\sqrt{pq})$, and slowing down the rate of convergence in (1).

In order to obtain the correct exponents of convergence, we need to **redo the coupling rules** as follows. We now let the movements of X_t and Y_t be synchronized whenever both are not at zero (i.e. $\{X_t, Y_t\} \cap \{0\} = \emptyset$), while letting X_t and Y_t move independently when one of them is at zero, and the other is not. Then at the hitting time τ , either $X_t = Y_t = 0$ and the processes are successfully coupled, or $X_t = 1$ and $Y_t = 0$. In the latter case we are back to the geometric variable with the failure probability of $\frac{q}{q+r}$. That is, the only way for X_t and Y_t to couple would be if one of the two is at state 0 and the other is at state 1. Using the set theory notations, if

$\{X_t, Y_t\} = \{0, 1\}$, conditioning on $\{X_{t+1}, Y_{t+1}\} \neq \{1, 2\}$ would give us

$$\{X_{t+1}, Y_{t+1}\} = \begin{cases} \{1\} & \text{with probability } \frac{r}{q+r}, \\ \{0, 1\} & \text{with probability } \frac{q}{q+r}, \end{cases}$$

When $\frac{q}{q+r} > r + 2\sqrt{pq}$, the above modified coupling captures the order $\left(\frac{q}{q+r}\right)^t$. The coefficient A however is much harder to estimate using the coupling approach, while it is immediately provided in Theorem 1 and its corollary. Take for example $p = \frac{1}{11}$, $r = \frac{1}{11}$ and $q = \frac{9}{11}$. There $\frac{q}{q+r} > r + 2\sqrt{pq}$, and according to the Corollary, the lower bound of $A \left(\frac{q}{q+r}\right)^t - B(r + 2\sqrt{pq})^t$ and the upper bound of $A \left(\frac{q}{q+r}\right)^t + B(r + 2\sqrt{pq})^t$ are of the matching order, and the order of convergence is tight

$$\|\nu - \mu_t\|_{TV} = \frac{91}{171} \left(\frac{9}{10}\right)^t \pm \frac{39}{28} \left(\frac{7}{11}\right)^t$$

2.2 Drift, minorization and geometric ergodicity

The optimal “energy function” $V(x) = \left(\frac{q}{p}\right)^{x/2}$ converts the geometric drift inequality in Meyn and Tweedie [6] Chapter 15 into equality

$$E[V(X_{t+1}) \mid X_t = x] = (r + 2\sqrt{pq})V(x) + \left(\sqrt{\frac{q}{p}} - (r + 2\sqrt{pq})\right) \mathbb{1}_C(x)$$

thus confirming the geometric convergence rate of $(r + 2\sqrt{pq})^t$ for the tail probability $P(\tau_C > t)$, where $C = \{0\}$ is the obvious choice for the “small set”, and τ_C is the hitting time. Once again all the challenge is at the origin. In fact there is only a trivial “minorization condition” when $C = \{0\}$. The minorization condition reads

$$p(x, A) \geq \epsilon Q(A) \quad \forall x \in C, \quad A \subset \Omega,$$

where, if $C = \{0\}$, the only choice for the probability measure Q is $Q = \delta_1$, and $\epsilon = 1$. With $\epsilon = 1$ the split of the Markov chain is trivial, and as far as the corresponding coupling goes, the only issue would be (as we mentioned before) to compute the tail of the hitting time $\min\{t : (X_t, Y_t) \in C \times C\}$ when q is large. If $C = \{0, 1, \dots, k\}$ for some $k > 0$, there is no minorization condition. In the latter case, estimating the hitting time $\min\{t : (X_t, Y_t) \in C \times C\}$ is straightforward, but without minorization, this will not be enough to estimate the tail for the coupling time. The “splitting technique” will not work, rather a coupling approach of the preceding subsection to be pursued.

The case of recurrent reflecting random walk (the M/M/1 queue) had been considered as one of the four benchmark examples in the geometric ergodicity theory (see [1] and references therein). There, in the absence of the second largest eigenvalue of $\left(-\frac{q}{q+r}\right)$, with $r = 0$, the rate of $(2\sqrt{pq})^t$ was proven to be the optimal (see [5]). The methods in the theory of geometric convergence are in the most part based on the renewal theory (see [6], [1], [7] and references therein), and concentrate more on the tail probability for the hitting time τ_C in the splitting method. As for the Markov chain P considered in this paper, in the absence of a useful splitting of it, the approach that works is the coupling. While the coupling provides the right exponents, it does not necessarily produce the tight coefficients.

3 The proof of Theorem 1

Proof. Since we require $r > 0$ for aperiodicity, we will need to obtain the spectral measure ψ via an argument similar to that of Karlin and McGregor in [2], where the case of $r = 0$ was solved. The orthogonal polynomials are obtained via solving a simple linear recursion: $Q_0 = 1$, $Q_1 = \lambda$, and $Q_n(\lambda) = c_1(\lambda)\rho_1^n(\lambda) + c_2(\lambda)\rho_2^n(\lambda)$, where $\rho_1(\lambda) = \frac{\lambda-r+\sqrt{(\lambda-r)^2-4pq}}{2p}$ and $\rho_2(\lambda) = \frac{\lambda-r-\sqrt{(\lambda-r)^2-4pq}}{2p}$ are the roots of the characteristic equation for the recursion, and $c_1 = \frac{\rho_2-\lambda}{\rho_2-\rho_1}$ and $c_2 = \frac{\lambda-\rho_1}{\rho_2-\rho_1}$.

Now $\pi_0 = 1$, $\pi_n = \frac{p^{n-1}}{q^n}$ ($n \geq 1$) and $\rho = \frac{q-p+1}{q-p}$. Also, we observe that

$$\begin{cases} |\rho_2(\lambda)| > \sqrt{\frac{q}{p}} & \text{on } [-1, r - 2\sqrt{pq}), \\ |\rho_2(\lambda)| < \sqrt{\frac{q}{p}} & \text{on } (r + 2\sqrt{pq}, 1], \\ |\rho_2(\lambda)| = \sqrt{\frac{q}{p}} & \text{on } [r - 2\sqrt{pq}, r + 2\sqrt{pq}], \end{cases}$$

and $\rho_1\rho_2 = \frac{q}{p}$.

The above will help us to identify the point mass locations in the measure ψ since each point mass in ψ occurs when $\sum_k \pi_k Q_k^2(\lambda) < \infty$. Thus we need to find all $\lambda \in (r + 2\sqrt{pq}, 1]$ such that $c_1(\lambda) = 0$ and all $\lambda \in [-1, r - 2\sqrt{pq})$ such that $c_2(\lambda) = 0$. There are two roots, $\lambda = 1$ and $\lambda = -\frac{q}{q+r}$.

We already know everything about the point mass at $\lambda = 1$: $Q_k(1) = 1$ for all $k \geq 0$, and $\rho = \sum_{k=0}^{\infty} \pi_k Q_k^2(1) = \frac{1+q-p}{q-p}$ is the reciprocal of the point mass at $\lambda = 1$.

The only other point mass is at $\lambda = -\frac{q}{q+r}$. One can verify that $\rho_1\left(-\frac{q}{q+r}\right) = -\frac{q}{q+r}$ and $Q_k\left(-\frac{q}{q+r}\right) = \left(-\frac{q}{q+r}\right)^k$, and therefore

$$\sum_{k=0}^{\infty} \pi_k Q_k^2\left(-\frac{q}{q+r}\right) = 1 + \frac{q}{(q+r)^2 - pq} = \frac{(1+q-p)(q+r)}{(1+q-p)(q+r) - q}$$

is the reciprocal of the point mass at $\lambda = -\frac{q}{q+r}$.

It follows that the rest of the mass of ψ (other than the two point masses) is spread inside $[r - 2\sqrt{pq}, r + 2\sqrt{pq}]$. In order to find the density of ψ inside $[r - 2\sqrt{pq}, r + 2\sqrt{pq}]$ we need to find $(e_0, (P - sI)^{-1}e_0)$ for $\text{Im}(s) \neq 0$, i.e. the upper left element in the resolvent of P . Let $(a_0(s), a_1(s), \dots)^T = (P - sI)^{-1}e_0$, then

$$-sa_0 + a_1 = 1, \quad \text{and} \quad qa_{n-1} + (r-s)a_n + pa_{n+1} = 0$$

Thus $a_n(s) = \alpha_1\rho_1(s)^n + \alpha_2\rho_2(s)^n$, where $\alpha_1 = \frac{a_0(\rho_2-s)-1}{\rho_2(s)-\rho_1(s)}$ and $\alpha_2 = \frac{1-a_0(\rho_1-s)}{\rho_2(s)-\rho_1(s)}$.

Since $(a_0, a_1, \dots) \in \ell^2(\mathbb{C}, \pi)$,

$$|a_n| \sqrt{\frac{q^n}{p^n}} \rightarrow 0 \quad \text{as} \quad n \rightarrow +\infty$$

Hence when $|\rho_1(s)| \neq |\rho_2(s)|$, either $\alpha_1 = 0$ or $\alpha_2 = 0$, and therefore

$$a_0(s) = \frac{\chi_{|\rho_1(s)| < \sqrt{\frac{q}{p}}}}{\rho_1(s) - s} + \frac{\chi_{|\rho_2(s)| < \sqrt{\frac{q}{p}}}}{\rho_2(s) - s} \quad (2)$$

Now, because of the point masses at 1 and $-\frac{q}{q+r}$, $a_0(s) = \int_{(-1,1]} \frac{d\psi(z)}{z-s}$ can be expressed as

$$a_0(s) = \frac{q-p}{1+q-p} \left(\frac{1}{1-s} \right) + \frac{(1+q-p)(q+r)-q}{(1+q-p)(q+r)} \left(\frac{1}{-\frac{q}{q+r}-s} \right) + \int_{(-1,1)} \frac{\varphi(z)dz}{z-s},$$

where $\varphi(z)$ is an atom-less function. Next we will use the following basic property of Cauchy transforms $Cf(s) = \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f(z)dz}{z-s}$ that can be derived using the Cauchy integral formula, or similarly, an approximation to the identity formula ¹:

$$C_+ - C_- = I \quad (3)$$

Here $C_+f(z) = \lim_{s \rightarrow z: \text{Im}(s) > 0} Cf(s)$ and $C_-f(z) = \lim_{s \rightarrow z: \text{Im}(s) < 0} Cf(s)$ for all $z \in \mathbb{R}$. The above equation (3) implies

$$\varphi(x) = \frac{1}{2\pi i} \left(\lim_{s=x+i\varepsilon: \varepsilon \rightarrow 0+} a_0(s) - \lim_{s=x-i\varepsilon: \varepsilon \rightarrow 0+} a_0(s) \right)$$

for all $x \in (-1,1)$. Recalling (2), we express φ as $\varphi(x) = \frac{\rho_1(x)-\rho_2(x)}{2\pi i(\rho_1(x)-x)(\rho_2(x)-x)}$ for $x \in (r-2\sqrt{pq}, r+2\sqrt{pq})$, which in turn simplifies to

$$\varphi(x) = \begin{cases} \frac{\sqrt{(x-r)^2-4pq}}{2\pi i((r+q)x+q)(1-x)} & \text{if } x \in (r-2\sqrt{pq}, r+2\sqrt{pq}), \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathcal{I} = (r-2\sqrt{pq}, r+2\sqrt{pq})$ denote the support interval, and let $\mathbb{1}_{\mathcal{I}}(x)$ be its indicator function. Here

$$\int_{-1}^1 \varphi(x)dx = \frac{p}{q+r}$$

and one can check that

$$\psi(x) = \frac{q-p}{1+q-p} \cdot \delta_1(x) + \frac{(1+q-p)(q+r)-q}{(1+q-p)(q+r)} \cdot \delta_{-\frac{q}{q+r}}(x) + \frac{\sqrt{4pq-(x-r)^2}}{2\pi((r+q)x+q)(1-x)} \cdot \mathbb{1}_{\mathcal{I}}(x)$$

integrates to one.

Observe that the residues of $g(z) = \frac{\sqrt{(z-r)^2-4pq}}{((r+q)z+q)(1-z)}$ are

$$\text{Res}(g(z), 1) = \frac{q-p}{1+q-p} \quad \text{and} \quad \text{Res}\left(g(z), -\frac{q}{q+r}\right) = \frac{(1+q-p)(q+r)-q}{(1+q-p)(q+r)}$$

in the principle branch of the log function.

¹The curve in the integral does not need to be \mathbb{R} for $C_+ - C_- = I$ to hold.

Now

$$\int_{(-1,1)} \lambda^t Q_n(\lambda) d\psi(\lambda) = \frac{(1+q-p)(q+r)-q}{(1+q-p)(q+r)} \cdot \left(-\frac{q}{q+r}\right)^{t+n} + \int_{r-2\sqrt{pq}}^{r+2\sqrt{pq}} \lambda^t (c_1 \rho_1^n + c_2 \rho_2^n) \frac{\rho_1 - \rho_2}{2\pi i (\rho_1 - \lambda)(\rho_2 - \lambda)} d\lambda$$

and therefore, since $c_1 = \frac{\rho_2 - \lambda}{\rho_2 - \rho_1}$ and $c_2 = \frac{\lambda - \rho_1}{\rho_2 - \rho_1}$,

$$\int_{(-1,1)} \lambda^t Q_n(\lambda) d\psi(\lambda) = \frac{(1+q-p)(q+r)-q}{(1+q-p)(q+r)} \cdot \left(-\frac{q}{q+r}\right)^{t+n} + \frac{1}{2\pi i} \int_{r-2\sqrt{pq}}^{r+2\sqrt{pq}} \lambda^t \left(\frac{\rho_2^n}{\rho_2 - \lambda} - \frac{\rho_1^n}{\rho_1 - \lambda} \right) d\lambda,$$

where, if we let $\rho_1 = \sqrt{\frac{q}{p}}z$ for z in the lower semicircle and $\rho_2 = \sqrt{\frac{q}{p}}z$ for z in the upper semicircle, then

$$\begin{aligned} \frac{1}{2\pi i} \int_{r-2\sqrt{pq}}^{r+2\sqrt{pq}} \lambda^t \left(\frac{\rho_2^n}{\rho_2 - \lambda} - \frac{\rho_1^n}{\rho_1 - \lambda} \right) d\lambda &= \left(\sqrt{\frac{q}{p}} \right)^n \frac{1}{2\pi i} \oint_{|z|=1} \frac{(\sqrt{pq}(z+z^{-1})+r)^t z^n \sqrt{pq}(1-z^{-2}) dz}{\sqrt{\frac{q}{p}}z - (\sqrt{pq}(z+z^{-1})+r)} \\ &= \left(\sqrt{\frac{q}{p}} \right)^n \left(\frac{p}{q+r} \right) \frac{1}{2\pi i} \oint_{|z|=1} \frac{(\sqrt{pq}(z+z^{-1})+r)^t z^n (z-z^{-1})}{\left(z - \sqrt{\frac{p}{q}} \frac{r+(1+q-p)}{2(q+r)} \right) \left(z - \sqrt{\frac{p}{q}} \frac{r-(1+q-p)}{2(q+r)} \right)} dz \end{aligned}$$

Here the absolute value of the function in the last integral is bounded by $M(r+2\sqrt{pq})^t$ with $M = \frac{2}{\left(1 - \sqrt{\frac{p}{q}} \frac{r+(1+q-p)}{2(q+r)}\right) \left(1 + \sqrt{\frac{p}{q}} \frac{r-(1+q-p)}{2(q+r)}\right)}$. Therefore, plugging in the values of π_n , we

show that the distance to stationarity $\|\nu - \mu_t\|_{TV} = \frac{1}{2} \sum_{n=0}^{\infty} \pi_n \left| \int_{(-1,1)} \lambda^t Q_n(\lambda) d\psi(\lambda) \right|$ is bounded above by

$$A \left(\frac{q}{q+r} \right)^t + B(r+2\sqrt{pq})^t$$

where

$$A = \frac{(1+q-p)(q+r)-q}{2(1+q-p)(q+r)} \sum_{n=0}^{\infty} \pi_n \left(\frac{q}{q+r} \right)^n = \frac{(1+q-p)(q+r)-q}{(1+q-p)(1-2p)}$$

and

$$B = \frac{M}{2} \left(\frac{p}{q+r} \right) \left(1 + \frac{1}{\sqrt{pq}-p} \right) = \frac{\left(\frac{p}{q+r} \right) \left(1 + \frac{1}{\sqrt{pq}-p} \right)}{\left(1 - \sqrt{\frac{p}{q}} \frac{r+(1+q-p)}{2(q+r)} \right) \left(1 + \sqrt{\frac{p}{q}} \frac{r-(1+q-p)}{2(q+r)} \right)}$$

The above upper bound can be improved if one obtains a better estimate of the trigonometric integrals involved in the sum.

We conclude that $t_{mix}(\varepsilon) = O\left(\frac{\log(\varepsilon)}{\log m(p,q)}\right)$ as $\varepsilon \downarrow 0$. □

Acknowledgment

The author would like to acknowledge useful comments about the idea of using orthogonal polynomials for computing mixing times he received from R.Burton, A.Dembo, P.Diaconis, M.Ossiander, E.Thomann, E.Waymire and J.Zuñiga.

References

- [1] P.H.Baxendale, *Renewal theory and computable convergence rates for geometrically ergodic Markov chains* Annals of Applied Probability, **15** No.1B, (2005), pp.700-738
- [2] S.Karlin and J.L.McGregor, *Random Walks* Illinois Journal of Math., **3**, No. 1, (1959), pp.417-431
- [3] Y.Kovchegov, *Orthogonality and probability: beyond nearest neighbor transitions* Electronic Communications in Probability, **14** (2009), pp.90-103
- [4] T.Lindvall, *Lectures on the coupling method* John Wiley & Sons, Inc., New York, (1992)
- [5] R.B.Lund and R.L.Tweedie, *Geometric convergence rates for stochastically ordered Markov chains* Math. Oper. Res., **21** (1996), pp.182-194
- [6] S.Meyn and R.L.Tweedie, *Markov Chains and Stochastic Stability* Second edition, Cambridge University Press, (2009)
- [7] J.S.Rosenthal, *Minorization conditions and convergence rates for Markov chain Monte Carlo* J. Amer. Stat. Assoc., **90**, (1995), pp.558-566
- [8] G.Szegő, *Orthogonal Polynomials* Fourth edition. AMS Colloquium Publications, Vol. **23**, (1975)